Disjointness is hard in the multiparty number-on-the-forehead model

Troy Lee
Department of Computer Science
Columbia University *

Adi Shraibman
Department of Mathematics
Weizmann Institute of Science †

Abstract

We show that disjointness requires randomized communication $\Omega\left(\frac{n^{1/(k+1)}}{2^{2^k}}\right)$ in the general k-party number-on-the-forehead model of complexity. The previous best lower bound for $k \geq 3$ was $\frac{\log n}{k-1}$. Our results give a separation between nondeterministic and randomized multiparty number-on-the-forehead communication complexity for up to $k = \log\log n - O(\log\log\log n)$ many players. Also by a reduction of Beame, Pitassi, and Segerlind, these results imply subexponential lower bounds on the size of proofs needed to refute certain unsatisfiable CNFs in a broad class of proof systems, including tree-like Lovász-Schrijver proofs.

1 Introduction

Since its introduction thirty years ago [Abe78, Yao79], communication complexity has become a key concept in complexity theory and theoretical computer science in general. Part of its appeal is that it has applications to many different computational models, for example to formula size and circuit depth, proof complexity, branching programs, VLSI design, and time-space trade-offs for Turing machines (see [KN97] for more details).

One area of communication complexity which still holds many mysteries is the k-party "number-on-the-forehead" model, originally introduced by Chandra, Furst, and Lipton [CFL83]. In this model, k parties wish to compute a function $f : (\{-1,+1\}^n)^k \to \{-1,1\}$. On input $(x_1,\ldots,x_k)$, the i^{th} player receives $(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_k)$. That is, player i has knowledge of the entire input *except* for the string x_i, which figuratively can be thought of as sitting on his forehead. The players communicate by writing messages "on a blackboard," so that all players see each message. The large overlap in the player's knowledge is part of what makes showing lower bounds in this model so difficult. This difficulty, however, is rewarded by the richness and strength of

*Work conducted at Rutgers University, supported by a NSF Mathematical Sciences Postdoctoral Fellowship. Email: troyjlee@gmail.com
†Email: adi.shraibman@weizmann.ac.il

consequences of such lower bounds: for example, by results of [HG91, BT94], showing a super-polylogarithmic lower bound on an explicit function for polylogarithmic many players would give an explicit function outside of the class ACC^0 — that is, a function which requires super-polynomial size constant-depth circuits using AND, OR, NOT, and modulo m gates.

While showing such bounds remains a challenging open problem, we do know of explicit functions which require large communication in this model for $\Theta(\log n)$ many players. Babai, Nisan, and Szegedy [BNS89] showed that the inner product function generalized to k-parties requires randomized communication $\Omega(n/4^k)$, and for other explicit functions slightly larger bounds of size $\Omega(n/2^k)$ are known [FG05]. These lower bounds are all achieved using the discrepancy method, a very general technique which gives lower bounds even on randomized models with error probability close to $1/2$, and also on nondeterministic communication complexity.

For some basic functions, however, there is a huge gap in our knowledge. One example is the disjointness function, or equivalently its complement, set intersection. In the set intersection problem, the goal of the players is to determine if there is an index j such that every string x_i has a -1 in position j, where here and throughout the paper we interpret -1 as 'true.' The best known protocol has cost $O(k^2 n \log(n)/2^k)$ [Gro94]. On the other hand, the best lower bound in the general number-on-the-forehead model is $\frac{\log n}{k-1}$, for $k \geq 3$ [Tes02, BPSW06]. For $k = 2$ tight bounds are known of $\Theta(n)$ for randomized communication complexity [KS87] and $\Theta(\sqrt{n})$ for quantum communication complexity [Raz03, AA05].

A major obstacle toward proving better lower bounds on set intersection is that it has a low cost nondeterministic protocol. In case there is a position where all players have a -1, with $O(\log n)$ bits a prover can send the name of this position and the players can then verify this is the case. Since the discrepancy method is also a lower bound on nondeterministic complexity, it is limited to logarithmic lower bounds for set intersection. Even in the two-party case, determining the complexity of set intersection in the randomized and quantum models was a long-standing open problem, in part for this reason.

In the multiparty case, the discrepancy method is the only technique which has been used to show lower bounds on the general randomized model of number-on-the-forehead complexity. Although other two-party methods can be generalized to the multiparty number-on-the-forehead model, they can become very difficult to handle. One source of this difficulty is that, whereas in the two party case we can nicely represent the function $f(x, y)$ as a matrix, in the multiparty case we deal with higher dimensional tensors. This makes many of the linear algebraic tools so useful in the two-party case inapplicable or at least much more involved. For example, while matrix rank is a staple lower bound technique for deterministic two-party complexity, in the tensor case even basic questions like the maximum rank of a $n \times n \times n$ tensor remain open.

Besides this technical challenge, additional motivation to studying the number-on-the-forehead complexity of disjointness was given by Beame, Pitassi, and Segerlind [BPS06], who showed that lower bounds on disjointness imply lower bounds on a very general class of proof systems, including cutting planes and Lovász-Schrijver proof systems.

We show that disjointness requires randomized communication $\Omega\left(\frac{n^{1/(k+1)}}{2^{2^k}}\right)$ in the general k-party number-on-the-forehead model. This separates nondeterministic and randomized multiparty number-on-the-forehead complexity for up to $k = \log \log n - O(\log \log \log n)$ many players. Also by the work of [BPS06] this implies subexponential lower bounds on the size of proofs needed to refute certain unsatisfiable formulas by tree-like proofs in Lovász-Schrijver and more powerful proof systems.

Chattopadhyay and Ada [CA08] have independently obtained similar bounds on disjointness using similar techniques.

1.1 Related work

For restricted models of computation, bounds are known which are stronger than ours. Wigderson showed that for one-way three-party number-on-the-forehead protocols, disjointness requires communication $\Omega(n^{1/2})$ (this result appears in [BHK01]). More recently, Viola and Wigderson [VW07] extended this approach to show a bound of $\Omega(n^{1/(k-1)}/k^{O(k)})$ on the complexity of one-way k-party protocols computing disjointness. These results actually show bounds on a pointer jumping function which reduces to disjointness.

Beame, Pitassi, Segerlind, and Wigderson [BPSW06] devised a method based on a direct product theorem to show a $\Omega(n^{1/3})$ bound on the complexity of three-party disjointness in a model stronger than one-way where the first player speaks once, and then the two remaining players interact arbitrarily.

Following up on our work, David, Pitassi, and Viola [DPV08] gave an explicit function which separates nondeterministic and randomized number-on-the-forehead communication complexity for up to $\Omega(\log n)$ players. They are also able, for any constant c to give a function computable in AC^0 which separates them for up to $c \log\log n$ players. Note that disjointness can be computed in AC^0, but that our bounds are already trivial for $\log\log n$ players. Even more recently, Beame and Huynh-Ngoc [BHN08] have shown a bound of $2^{\Omega(\sqrt{\log n}/\sqrt{k})-k}$ on the k-party number-on-the-forehead complexity of disjointness. This bound remains non-trivial for up to $\Theta(\log^{1/3} n)$ many players, but is not as strong as our bound for few players.

1.2 Overview of techniques

There is a natural correspondence between functions $f : (\{-1,+1\}^n)^k \to \{-1,1\}$ and sign k-tensors. Sometimes it is more convenient to consider the function form, and sometimes, like when discussing norms, it is more convenient to consider tensors.

Our proof combines two ingredients. The first of these is the notion of an approximation norm. For a norm Φ, and a sign tensor A, the *approximation norm* associated to Φ and A, denoted $\Phi^\alpha(A)$, is the smallest Φ norm of an element 'close' to A. Here α quantifies the term 'close.'

Approximation norms turn out to be quite useful for showing lower bounds on randomized and quantum communication complexity [Kla01, Raz03, LS07]. Razborov, for example, uses the approximation trace norm to prove a tight lower bound on the quantum communication complexity of set intersection.

We use what we call the *cylinder intersection norm*, denoted μ. This norm can be seen as a multiparty generalization of a quantity used in Lemma 3.1 of Klauck [Kla01]. As a correct deterministic protocol partitions the communication matrix into rectangles on which the function is constant, analogously a correct deterministic number-on-the-forehead protocol decomposes the communication tensor into cylinder intersections on which the function is constant. Roughly speaking, $\mu(A)$ measures how efficiently A can be written as a sum of cylinder intersections. In this way, if A has low communication complexity, it will also have low μ norm. We defer formal definitions to Section 3.

We denote the approximate version of the cylinder intersection norm by μ^α where $1 \leq \alpha < \infty$ represents the measure of approximation. This measure provides a lower bound on randomized communication complexity in the number-on-the-forehead model. The limiting case $\mu^\infty(A)$ turns out to be exactly the usual discrepancy method. For bounded α we obtain a technique which is strictly stronger than the discrepancy method.

Following [LMSS07, LS07], to show lower bounds on $\mu^\alpha(A)$, we write it in terms of the dual norm μ^*. By definition of a dual norm, we have

$$\mu(B) = \max_Q \frac{\langle B, Q \rangle}{\mu^*(Q)}. \tag{1}$$

This "max" formulation of μ is often more convenient for showing lower bounds. The dual norm μ^* is closely related to discrepancy with respect to the uniform distribution, so we can use existing techniques to upper bound $\mu^*(Q)$.

This formulation of μ also gives a way to write μ^α in terms of a maximization quantity.

$$\mu^\alpha(A) = \max_Q \frac{(1+\alpha)|\langle A, Q \rangle| + (1-\alpha)\|Q\|_1}{2\mu^*(Q)}. \tag{2}$$

All one needs for showing lower bounds is that the left hand side is at least as large as the right hand side. This can be shown quite simply using Equation 1 and elementary inequalities and was noted, for example, by Razborov in the context of the approximation trace norm. The fact that equality holds here requires the use of linear programming duality or a separation theorem for convex bodies and seems to be less well known.

As the dual norm μ^* is essentially discrepancy with respect to the uniform distribution, the approximation μ norm can be seen as an extension of discrepancy in another way. Instead of proving that the tensor of interest A has small discrepancy, it is enough to prove that there is a tensor Q which has small discrepancy and has large correlation with A, relative to $\|Q\|_1$. This is why this method is called *generalized discrepancy* in [CA08].

To find a good witness tensor Q, we use ideas from a second line of research. While the norm framework of Equation (2) provides a nice approach to lower bound communication complexity, it gives no hint about how to choose a good witness Q—in general a difficult problem. Works by Sherstov [She07, She08] and Shi and Zhu [SZ07] in the two-party case, and Chattopadhyay [Cha07] in the multiparty case provide an elegant way to choose a good witness for a general class of matrices and tensors. These works look at block composed functions of the form $f \circ g^n(x_1, \ldots, x_k) = f(g(x_1^1, \ldots, x_k^1), \ldots, g(x_1^n, \ldots, x_k^n))$. Notice that set intersection is a block composed function where $f = \mathrm{OR}_n$ is the OR function on n bits and $g = \mathrm{AND}_k$ is the k-player AND function on one bit. Sherstov [She07] first showed that when $g(x, i) = x_i$, the discrepancy of a block composed function could be bounded in terms of the threshold degree of f, the minimum degree of a polynomial which agrees in sign with f on the Boolean cube. Building on this result, Chattopadhyay showed an analogous statement in the number-on-the-forehead case for an appropriately generalized multiparty function g.

Sherstov and independently Shi-Zhu showed that the approximate trace norm of a block composed function could be lower bounded in terms of the approximate degree of f, again provided that the inner function g satisfies certain technical conditions. The μ norm provides bounds at least as large as the trace norm method [LS07], thus these works also lower bound μ^α. In this paper, we take the natural step to show that μ^α of a block composed multiparty function can be lower bounded in terms of the approximate degree of f, for a particular multiparty inner function g such that the composed function $f \circ g^n$ can be embedded in the set intersection problem.

1.3 Consequences for Lovász-Schrijver proof systems and beyond

Beame, Pitassi, and Segerlind [BPS06] show that bounds on multiparty disjointness imply strong lower bounds on the size of refutations of certain unsatisfiable formulas, for a very general class of proof systems. We now introduce and motivate the study of these proof systems. Formal definitions and the implications of our results will be given in Section 6.2.

The fact that linear and semidefinite programs can be solved with high precision in polynomial time is a remarkable algorithmic achievment. It is thus interesting to ask how these algorithms fare when pitted against NP-complete problems. For many NP-complete problems, there is a very natural approach to solving them via linear or semidefinite programming: namely, we first formulate the problem as optimizing a convex function over the Boolean cube, i.e. with variables subject to the quadratic constraints $x_i^2 = x_i$. We then relax these quadratic constraints to linear or semidefinite constraints to obtain a program which can be solved in polynomial time. For example, a linear relaxation of $x_i^2 = x_i$ may simply be the constraint $0 \leq x_i \leq 1$. In the case of vertex cover, for example, such a simple relaxation already gives a linear program with approximation ratio of 2. Semidefinite constraints are in general more complicated, but there are several "automatic" ways of generating valid semidefinite inequalities—that is, semidefinite inequalities satisfied by all Boolean solutions of the original problem. Perhaps the best known of these is the Lovász-Schrijver "lift and project" method [LS91]. The seminal 0.878-approximation algorithm for MAXCUT of Goemans and Williamson [GW95] can be obtained by relaxing the natural Boolean programming problem with semidefinite constraints obtained by one application of the Lovász-Schrijver method.

As these techniques have given impressive results in approximation algorithms, it is natural to ask if they can also be used to efficiently obtain exact solutions. Namely, how many inequalities need to be added in general until all fractional optima are eliminated and only true Boolean optima remain?

One way to address this question is to consider proof systems with derivation rules based on linear programming or the Lovász-Schrijver method. Our particular application will look at the size of proofs needed to refute unsatisfiable formulas. Given a CNF ϕ, we can naturally represent the satisfiability of ϕ as the satisfiability of a system of linear inequalities, one for each clause. For example, the clause $x_1 \vee x_4 \vee \neg x_5$ would be represented as $x_1 + x_4 + (1 - x_5) \geq 1$. Suppose that ϕ is unsatisfiable. Then consider a proof system in which the "axioms" are the inequalities obtained from the clauses of ϕ, and the goal is to derive the contradiction $0 \geq 1$. By the results of [BPS06], our results on disjointness imply that there are unsatisfiable formulas such that any refutation obtained by generating new inequalities by the Lovász-Schrijver method in a "tree-like" way requires size $2^{n^{\Omega(1)}}$. For a standard formulation of the Lovász-Schrijver method known as LS_+, bounds of size $2^{\Omega(n)}$ for tree-like proofs have already been shown by very different methods [IK06].

The advantage of the number-on-the-forehead communication complexity approach, however, is that it can also be applied to much more powerful proof systems which are currently untouchable by other methods. Beame, Pitassi, and Segerlind [BPS06] show that lower bounds on k-party communication complexity of disjointness give lower bounds on the size of tree-like proofs of certain unsatisfiable CNFs $\phi(x)$, where the derivation rule is as follows: from inequalities f, g of degree $k-1$ in x, we are allowed to conclude a degree $k-1$ inequality h if every Boolean assignment to x which satisfies f and g also satisfies h. Lovász-Schrijver proof systems are a special case of such degree-2 systems. Our bounds on disjointness imply the existence of unsatisfiable formulas whose refutation requires subexponential size tree-like degree-k proofs, for any constant k. [1] The aforementioned lower bounds on LS_+ proof systems strongly rely on specific properties of the Lovász-Schrijver operator—showing superpolynomial bounds on the size of tree-like proofs in the more general degree-k model was previously open even in the case $k = 2$.

2 Preliminaries and notation

We let $[n] = \{1, \ldots, n\}$. For multiparty communication complexity it is convenient to work with tensors, the generalization of matrices to higher dimensions. If an element of a tensor A is specified by k indices, we say that A is a k-tensor. For a k-tensor A of dimensions $(n_1, \ldots, n_k)$ we write $\mathrm{size}(A) = n_1 \cdots n_k$. A tensor for which all entries are in $\{-1, 1\}$ we call a sign tensor. For a function $f : X_1 \times \ldots \times X_k \to \{-1, 1\}$, we define the communication tensor corresponding to f to be a k-tensor A_f where $A_f[x_1, \ldots, x_k] = f(x_1, \ldots, x_k)$. We identify f with its communication tensor. For a set $Z \subseteq X_1 \times \ldots \times X_k$ we let $\chi(Z)$ be its characteristic tensor where $\chi(Z)[x_1, \ldots, x_k] = 1$ if $(x_1, \ldots, x_k) \in Z$ and is 0 otherwise.

For a sign tensor A, we denote by $D^k(A)$ the deterministic communication complexity of A in the k-party number-on-the-forehead model. The public coin randomized communication complexity with error bound $\epsilon \geq 0$ is denoted $R^k_\epsilon(A)$. We drop the superscript when the number of players is clear from context.

We use the shorthand $A \geq c$ to indicate that all of the entries of A are at least c. The Hadamard or entrywise product of two tensors A and B is denoted by $A \circ B$. Their inner product is denoted $\langle A, B \rangle = \sum_{x_1,\ldots,x_k} A[x_1, \ldots, x_k] B[x_1, \ldots, x_k]$. The ℓ_1 and ℓ_∞ norms of a tensor A are $\|A\|_1 = \sum_{x_1,\ldots,x_k} |A[x_1, \ldots, x_k]|$ and $\|A\|_\infty = \max_{x_1,\ldots,x_k} |A[x_1, \ldots, x_k]|$, respectively.

We also need some basic elements of Fourier analysis. For $S \subseteq [n]$ we define $\chi_S : \{-1, +1\}^n \to \{-1, 1\}$ as $\chi_S(x) = \prod_{i \in S} x_i$. As the χ_S form an orthogonal basis, for any function $f : \{-1, +1\}^n \to$

[1] The conference version of this paper reported bounds on degree-k proof systems for up to $k = \log\log n - O(\log\log\log n)$. As pointed out to us by Paul Beame, however, this is not justified by the reduction of [BPS06], which requires certain constraints on the size of k.

$\mathbb{R}$ we have a unique representation

$$f(x) = \sum_{S \subseteq n} \hat{f}(S)\chi_S(x)$$

where $\hat{f}(S) = (1/2^n)\langle f, \chi_S \rangle$, are the Fourier coefficients of f. The degree of f is the size of the largest set S for which $\hat{f}(S)$ is nonzero.

3 The Method

In this section we present a method for proving lower bounds on randomized communication complexity in the number-on-the-forehead model that generalizes and significantly strengthens the discrepancy method.

3.1 Cylinder intersection norm

In two-party communication complexity, a key role is played by combinatorial rectangles—subsets of the form $Z_1 \times Z_2$ where Z_1 is a subset of inputs to Alice and Z_2 is a subset of inputs to Bob. The analogous concept in the number-on-the-forehead model of multiparty communication complexity is that of a cylinder intersection.

Definition 1 (Cylinder intersection) *A subset $Z_i \subseteq X_1 \times \ldots \times X_k$ is called a cylinder in the i^{th} dimension if membership in Z_i does not depend on the i^{th} coordinate. That is, for every $(z_1, \ldots, z_i, \ldots, z_k) \in Z_i$ and $z'_i \in X_i$ it also holds that $(z_1, \ldots, z'_i, \ldots, z_k) \in Z_i$. A set Z is called a cylinder intersection if it can be expressed as $Z = \cap_{i=1}^{k} Z_i$ where each Z_i is a cylinder in the i^{th} dimension.*

Cylinder intersections are important because a correct deterministic number-on-the-forehead protocol for a function f partitions the corresponding communication tensor into cylinder intersections, each of which is monochromatic with respect to the function f.

Fact 2 *Let A be a sign k-tensor, and suppose that $D^k(A) \leq c$. Then there are cylinder intersections $Z_1, \ldots, Z_{2^c}$ such that*

$$A = \sum_{i=1}^{2^c} \alpha_i \chi(Z_i)$$

where $\alpha_i \in \{-1, +1\}$.

Our main object of study, termed the cylinder intersection norm, relaxes this notion of decomposition to allow $\alpha_i \in \mathbb{R}$. A similar such relaxation is done by [KKN95] in the context of nondeterministic communication complexity.

Cylinder intersection norm We denote by μ the norm induced by the absolute convex hull of the characteristic functions of all cylinder intersections. That is, for a k-tensor B

$$\mu(B) = \min\left\{\sum_i |\alpha_i| : B = \sum_i \alpha_i \chi(Z_i), \alpha_i \in \mathbb{R}\right\}$$

where each Z_i is a cylinder intersection and $\chi(Z_i)$ is its characteristic tensor.

In the two dimensional case, μ is very closely related to the γ_2 norm [LMSS07, LS07]. Indeed, for matrices B we have $\mu(B) = \Theta(\gamma_2(B))$.

Remark 3 *In our definition of μ above we chose to take $\chi(Z_i)$ as $\{0,1\}$ tensors. One can alternatively take them to be ± 1 valued tensors—a form which is sometimes easier to bound—without changing much. One can show*

$$\mu(B) \geq \mu_{\pm 1}(B) \geq 2^{-k}\mu(B).$$

where B is a k-tensor and $\mu_{\pm 1}(B)$ is defined as above with $\chi(Z_i)$ taking values from $\{-1,1\}$. In the matrix case, $\mu_{\pm}$ is also known as the nuclear norm [Jam87].

By Fact 2 we have the following.

Theorem 4 *It holds that $D^k(A) \geq \log(\mu(A))$ for every sign k-tensor A.*

A public coin randomized protocol is simply a probability distribution over deterministic protocols. This gives us the following fact:

Fact 5 *A sign k-tensor A satisfies $R_\epsilon^k(A) \leq c$ if and only if there are sign k-tensors A_i' for $i = 1,\dots,\ell$ satisfying $D^k(A_i') \leq c$ and a probability distribution $(p_1,\dots,p_\ell)$ such that*

$$\|A - \sum_{i=1}^{\ell} p_i A_i'\|_\infty \leq 2\epsilon.$$

To lower bound randomized communication complexity we consider an approximate variant of the cylinder intersection norm.

Definition 6 (Approximate cylinder intersection norm) *Let A be a sign k-tensor, and $\alpha \geq 1$. We define the α-approximate cylinder intersection norm as*

$$\mu^\alpha(A) = \min_B\{\mu(B) : 1 \leq A \circ B \leq \alpha\}$$

In words, we take the minimum of the cylinder intersection norm over all tensors B which are signed as A and have entries with magnitude between 1 and α. Considering the limiting case as $\alpha \to \infty$ motivates the definition

$$\mu^\infty(A) = \min_B\{\mu(B) : 1 \leq A \circ B\}$$

One should note that $\mu^\alpha(A) \leq \mu^\beta(A)$ for $1 \leq \beta \leq \alpha$.

The following theorem is an immediate consequence of the definition of the approximate cylinder intersection norm and Fact 5.

Theorem 7 *Let A be a sign k-tensor, and $0 \leq \epsilon < 1/2$. Then*

$$R_\epsilon^k(A) \geq \log(\mu^\alpha(A)) - \log(\alpha_\epsilon)$$

where $\alpha_\epsilon = 1/(1-2\epsilon)$ and $\alpha \geq \alpha_\epsilon$.

Proof: Let p_i and A'_i for $1 \leq i \leq \ell$ be as in Fact 5. We take

$$B = \frac{1}{1-2\epsilon} \sum_{i=1}^{\ell} p_i A'_i.$$

Notice that $1 \leq B \circ A \leq \alpha_\epsilon$, and hence by Definition 6

$$\mu^{\alpha_\epsilon}(A) \leq \mu(B).$$

Employing the fact that μ is a norm and Theorem 4, we get

$$\begin{aligned} \mu(B) &\leq \frac{1}{1-2\epsilon} \sum_i p_i \mu(A'_i) \\ &\leq \frac{1}{1-2\epsilon} \sum_i p_i 2^{D^k(A'_i)} \\ &\leq \frac{2^{R_\epsilon^k(A)}}{1-2\epsilon}. \end{aligned}$$

□

The nondeterministic complexity of a sign k-tensor A, denoted $N^k(A)$, is the logarithm of the minimum cardinality of a set of cylinder intersections $\{Z_i\}$ such that every entry of A with value -1 is covered by some Z_i, and no entry of A with value 1 is covered by Z_i. Notice that if $\{Z_i\}$ is such a covering of A, then letting $B = -\sum \chi(Z_i)$ we have $1 \leq A \circ (2B + J) < \infty$ where J is the all one tensor. As J is itself a cylinder, we have $\mu(J) = 1$, which gives the following.

Theorem 8 (folklore) *For a sign k-tensor A,*

$$N^k(A) \geq \log \frac{\mu^\infty(A) - 1}{2}$$

As we shall see in Section 3.3, μ^∞ is exactly the discrepancy method, which explains why the discrepancy method cannot show good lower bounds on disjointness, or indeed any function with low nondeterministic or co-nondeterministic communication complexity.

3.2 Employing duality

We now have a quantity, $\mu^\alpha(A)$, which can be used to prove lower bounds on randomized communication complexity in the number-on-the-forehead model. As this quantity is defined in terms of a minimization, however, it seems in itself a difficult quantity to bound from below.

In this section, we employ the duality theory of linear programming to find an equivalent formulation of $\mu^\alpha(A)$ in terms of a maximization problem. This makes the task of proving lower bounds for $\mu^\alpha(A)$ much easier, as the $\forall$ quantifier we had to deal with before is now replaced by an $\exists$ quantifier.

As it turns out, in order to prove lower bounds on $\mu^\alpha(A)$ we will need to understand the dual norm of μ, denoted μ^*. The standard definition of a dual norm is

$$\mu^*(Q) = \max_{B:\mu(B)\leq 1} |\langle B, Q\rangle|,$$

for any tensor Q. Since the unit ball of μ is the absolute convex hull of the characteristic vectors of cylinder intersections, we can alternatively write

$$\mu^*(Q) = \max_Z |\langle Q, \chi(Z)\rangle|$$

where the maximum is taken over all cylinder intersections Z.

It is instructive to compare this with the definition of discrepancy.

Definition 9 (discrepancy) *Let A be a sign k-tensor, and let P be a probability distribution on its entries. The discrepancy of A with respect to P, written $\mathrm{disc}_P(A)$, is*

$$\mathrm{disc}_P(A) = \max_Z |\langle A \circ P, \chi(Z)\rangle|$$

where the maximum is taken over cylinder intersections Z.

Thus we see that $\mathrm{disc}_P(A) = \mu^*(A \circ P)$, and we can use existing techniques for discrepancy to also upper bound μ^*.

As the dual of a dual norm is again the norm, we can write the μ norm as

$$\mu(B) = \max_Q \frac{\langle B, Q\rangle}{\mu^*(Q)}. \tag{3}$$

To prove our lower bounds, we will use an equivalent formulation of μ^α in terms of the dual norm μ^*.

Theorem 10 *Let A be a sign tensor and $1 \leq \alpha < \infty$.*

$$\mu^\alpha(A) = \max_Q \frac{(1+\alpha)\langle A, Q\rangle + (1-\alpha)\|Q\|_1}{2\mu^*(Q)}$$

When $\alpha = \infty$ we have

$$\mu^\infty(A) = \max_{Q: A\circ Q \geq 0} \frac{\langle A, Q\rangle}{\mu^*(Q)}$$

Proof: We can quite easily see that the left hand side is at least as large as the right hand side, which is all that is needed for proving lower bounds. By Equation (3) and the definition of μ^α we have

$$\mu^\alpha(A) = \min_{B:1 \le A \circ B \le \alpha} \max_Q \frac{\langle B, Q \rangle}{\mu^*(Q)}.$$

If we rewrite Q as the sum of two parts, Q^+, satisfying $Q^+ \circ A \ge 0$ and Q^- satisfying $Q^- \circ A < 0$ then we can see that

$$\mu^\alpha(A) \ge \max_{Q^+, Q^-} \frac{\langle A, Q^+ \rangle + \alpha \langle A, Q^- \rangle}{\mu^*(Q^+ + Q^-)}$$

It is now straightforward to verify that this expression can be reworked into the form given above in the two cases $1 \le \alpha < \infty$ and $\alpha = \infty$.

To see that this inequality holds with equality, we write μ^α as a linear program and then use duality to derive the dual expression given in the theorem. As it is easy to check that the primal program is feasible with a finite optimum, by Slater's condition these primal and dual forms coincide with the same finite value.

We treat the case $1 \le \alpha < \infty$ first. We can write $\mu^\alpha(A)$ as a linear program as follows. For each cylinder intersection Z_i let $X_i = \chi(Z_i)$. Then

$$\begin{aligned} \mu^\alpha(A) \quad &= \quad \min_{p,q} \sum_i p_i + q_i \\ \text{s.t.} \quad & \quad 1 \le \left(\sum_i (p_i - q_i) X_i \right) \circ A \le \alpha \\ & \quad p_i, q_i \ge 0 \end{aligned}$$

Taking the dual of this program in the straightforward way, we obtain

$$\begin{aligned} \mu^\alpha(A) \quad &= \quad \max_Q \frac{(1+\alpha)\langle A, Q \rangle + (1-\alpha)\|Q\|_1}{2} \\ \text{s.t.} \quad & \quad |\langle X_i, Q \rangle| \le 1, \text{ for all } X_i \end{aligned}$$

For $\alpha = \infty$ we get the same program as above without the constraint $(\sum_i (p_i - q_i) X_i) \circ A \le \alpha$. Dualizing this program gives the desired result. □

Let us take a moment to compare our approach with that of Chattopadhyay and Ada. They also use the approximation μ norm, but with an additive approximation factor rather than a multiplicative factor as we use. More precisely, they use the measure $\mu^\epsilon(A) = \min_{B:\|A-B\|_\infty \le \epsilon} \mu(B)$. The dual form of this measure has the form

$$\mu^\epsilon(A) = \max_Q \frac{\langle A, Q \rangle - \epsilon \|Q\|_1}{\mu^*(Q)}.$$

Chattopadhyay and Ada directly derive that this dual expression is a lower bound on multiparty *distributional* communication complexity. Yao's characterization of randomized complexity in terms of distributional complexity [Yao83] then gives that it is also a lower bound on randomized communication complexity. They do not mention the primal definition of μ^α, but other than that, their proof is similar in structure to ours. For our proof we do not use Yao's principle but apply duality directly on the measure μ rather than on the complexity class itself.

While our presentation through the primal version of the μ norm is perhaps not as familiar as that via distributional complexity, we feel it does have advantages. First of all, this discussion holds quite generally: for any norm Φ one can show using the separation theorem that the approximation version Φ^α has a dual characterization analogous to that in Theorem 10. Second, we feel that the primal definition of μ^α arises very naturally and gives insight into the origin of the dual formulation—we do not have to guess this formula but can derive it. Finally, it is interesting to note that the primal and dual formulations are *equivalent*. This means that we do not lose anything in considering the more convenient dual formulation for proving lower bounds.

3.3 The discrepancy method

Virtually all lower bounds in the general number-on-the-forehead model have used the discrepancy method. Let A be a sign tensor, and recall the definition of $\mathrm{disc}_P(A)$ from Section 3.2. Let $\mathrm{disc}(A) = \min_P \mathrm{disc}_P(A)$, where the minimum is taken over all probability distributions P. The discrepancy method turns out to be equivalent to $\mu^\infty(A)$.

Theorem 11

$$\mu^\infty(A) = \frac{1}{\mathrm{disc}(A)}.$$

Proof: By Theorem 10, for every sign tensor A

$$\mu^\infty(A) = \max_{Q \circ A \geq 0} \{\langle A, Q\rangle : \mu^*(Q) \leq 1\}$$

We can rewrite this as

$$\mu^\infty(A) = \max_{Q \circ A \geq 0} \frac{\langle A, Q\rangle}{\mu^*(Q)} = \max_{P : P \geq 0} \frac{\langle A, A \circ P\rangle}{\mu^*(A \circ P)}$$

As both numerator and denominator are homogeneous, we have

$$\begin{aligned}\mu^\infty(A) &= \max_{\substack{P:P\geq 0\\ \|P\|_1=1}} \frac{\langle A, A \circ P\rangle}{\mu^*(A \circ P)} = \max_{\substack{P:P\geq 0\\ \|P\|_1=1}} \frac{1}{\mu^*(A \circ P)} \\ &= \frac{1}{\mathrm{disc}(A)}.\end{aligned}$$

□

4 Techniques to bound $\mu^*(Q)$

In the last section, we saw that to bound the randomized number-on-the-forehead communication complexity of a sign tensor A, it suffices to find a tensor Q such that $\langle A, Q\rangle$ is large and $\mu^*(Q)$ is small. The first quantity is relatively simple and is in general not too hard to compute. Upper bounding $\mu^*(Q)$ is more subtle. In this section, we review some techniques for doing this.

In upper bounding the magnitude of the largest eigenvalue of a matrix B, a common thing is to consider the matrix BB^T, and use the fact that $\|B\|^2 \leq \|BB^T\|$. We will try to do a similar thing in upper bounding μ^*. In analogy with BB^T we make the next definition. Here and in what follows all expectations are taken with respect to the uniform distribution.

Definition 12 (Contraction product) *Let B be a k-tensor with entries indexed by elements from $X_1 \times \ldots \times X_k$. We define the contraction product of B along X_1, denoted $B \bullet_1 B$, to be a $2(k-1)$-tensor with entries indexed by elements from $X_2 \times X_2 \times \ldots \times X_k \times X_k$. The $x_2, x_2', \ldots, x_k, x_k'$ entry is defined to be*

$$B \bullet_1 B[x_2, x_2', \ldots, x_k, x_k'] = \mathbb{E}_{x_1}\left[\prod_{y_2 \in \{x_2, x_2'\}, \ldots, y_k \in \{x_k, x_k'\}} B[x_1, y_2, \ldots, y_k]\right]$$

The contraction product may be defined along other dimensions mutatis mutandis.

Notice that when B is a m-by-n matrix $B \bullet_1 B$ corresponds to $(1/m)BB^T$. In analogy with the fact that $\|B\|^2 \leq m\|B \bullet_1 B\|$, the next lemma gives a corresponding statement for the μ^* norm and k-tensors. This lemma originated in the work of Babai, Nisan, and Szegedy [BNS89] (see also [Chu90, Raz00]) and all lower bounds on the general model of randomized number-on-the-forehead complexity use some version of this lemma. The particular statement we use is from Chattopadhyay [Cha07].

Lemma 13 *Let B be a k-tensor. Then*

$$\left(\frac{\mu^*(B)}{\mathrm{size}(B)}\right)^{2^{k-1}} \leq \frac{\mu^*(B \bullet_1 B)}{\mathrm{size}(B \bullet_1 B)} \leq \mathbb{E}[|B \bullet_1 B|]$$

Proof: The second inequality follows since $\mu^*(X) \leq \|X\|_1$ for any real tensor X. The first inequality is standard, and follows by applying the Cauchy-Schwarz inequality repeatedly $k-1$ times. $\square$

4.1 Example: Hadamard tensors

We give an example to show how Lemma 13 can be used in conjunction with our μ method. Let H be a N-by-N Hadamard matrix. We show that $\mu^\infty(H) \geq \sqrt{N}$. Indeed, simply let the witness matrix Q be H itself. Incidentally, this corresponds to taking the uniform probability distribution in the discrepancy method. With this choice we clearly have $H \circ Q \geq 0$, and so

$$\mu^\infty(H) \geq \frac{\langle H, H \rangle}{\mu^*(H)} = \frac{N^2}{\mu^*(H)}$$

Now we bound $\mu^*(H)$ using Lemma 13 which gives:

$$\mu^*(H)^2 \leq N^4 \, \mathbb{E}[|H \bullet_1 H|] = N^3$$

as $H \bullet_1 H$ has nonzero entries only on the diagonal, and these entries are of magnitude one.

Ford and Gál [FG05] extend the notion of matrix orthogonality to tensors, defining what they call Hadamard tensors.

Definition 14 (Hadamard tensor) *Let H be a sign k-tensor. We say that H is a Hadamard tensor if*

$$(H \bullet_1 H)[x_2, x_2', \ldots, x_k, x_k'] = 0$$

whenever $x_i \neq x_i'$ for all $i = 2, \ldots, k$.

The simple proof above for Hadamard matrices can be easily extended to Hadamard tensors:

Theorem 15 (Ford and Gál [FG05]) *Let H be a Hadamard k-tensor of side length N. Then*

$$\mu^\infty(H) \geq \left(\frac{N}{k-1} \right)^{1/2^{k-1}}$$

Proof: We again take the witness Q to be H itself. This clearly satisfies $H \circ Q \geq 0$, and so

$$\mu^\infty(H) \geq \frac{\langle H, H \rangle}{\mu^*(H)} = \frac{N^k}{\mu^*(H)}$$

It now remains to upper bound $\mu^*(H)$ which we do by Lemma 13. This gives us

$$\mu^*(H)^{2^{k-1}} \leq N^{k2^{k-1}} \, \mathbb{E}[|H \bullet_1 H|]$$

The "Hadamard" property of H lets us easily upper bound $\mathbb{E}[|H \bullet_1 H|]$. Note that each entry of $H \bullet_1 H$ is of magnitude at most one, and the probability of a non-zero entry is at most

$$\Pr[\vee_{i=2}^k (x_i = x_i')] \leq \frac{k-1}{N}$$

by a union bound. Hence, we obtain

$$\mu^*(H)^{2^{k-1}} \leq (k-1) \frac{N^{k2^{k-1}}}{N}.$$

Putting everything together, we have

$$\mu^\infty(H) \geq \left(\frac{N}{k-1} \right)^{1/2^{k-1}}$$

□

Remark 16 *By doing a more careful inductive analysis, Ford and Gál obtain this result without the $k-1$ term in the denominator. They also construct explicit examples of Hadamard tensors.*

5 Lower bounds on μ^α for pattern tensors

In Section 5.1 we describe a key lemma which relates the approximate polynomial degree of f to the existence of a hard input "distribution" for f. This will only truly correspond to a distribution in the case of discrepancy—otherwise it can take on negative values. This lemma was first used in the context of communication complexity by Sherstov [She08] and independently by Shi and Zhu [SZ07].

In Section 5.2 we use this distribution, together with the machinery developed in Section 4 to prove lower bounds on a special kind of tensors, named pattern tensors. The application to disjointness appears in Section 6.1.

5.1 Dual polynomials

We define approximate degree in a slightly non-standard way to more smoothly handle both the bounded α and $\alpha = \infty$ cases.

Definition 17 *Let $f : \{-1,+1\}^n \to \{-1,1\}$. For $\alpha \geq 1$ we say that a function g gives an α-approximation to f if $1 \leq g(x)f(x) \leq \alpha$ for all $x \in \{-1,+1\}^n$. Similarly we say that g gives an ∞-approximation to f if $1 \leq g(x)f(x)$ for all $x \in \{-1,+1\}^n$. We let the α-approximate degree of f, denoted $\deg_\alpha(f)$, be the smallest degree of a function g which gives an α-approximation to f.*

Remark 18 *In a more standard scenario, one is considering a 0/1 valued function f and defines the approximate degree as $\deg'_\epsilon(f) = \min\{\deg(g) : \|f - g\|_\infty \leq \epsilon\}$. Letting $f_\pm$ be the sign representation of f, one can see that for $0 \leq \epsilon < 1/2$ our definition is equivalent to the standard one in the following sense: $\deg'_\epsilon(f) = \deg_{\alpha_\epsilon}(f_\pm)$ where $\alpha_\epsilon = \frac{1+2\epsilon}{1-2\epsilon}$.*

For a fixed natural number d, let $\alpha_d(f)$ be the smallest value of α for which there is a degree d polynomial which gives an α-approximation to f. Notice that $\alpha_d(f)$ can be written as a linear program. Namely, let $B(n,d) = \sum_{i=0}^{d} \binom{n}{i}$, and W be a 2^n-by-$B(n,d)$ incidence matrix, with rows labelled by strings $x \in \{-1,+1\}^n$ and columns labeled by monomials of degree at most d. We set $W(x,m) = m(x)$, where $m(x)$ is the evaluation of the monomial m on input x. Then

$$\alpha_d(f) = \min_y \{\|Wy\|_\infty : 1 \leq Wy \circ f\}$$

If this program is infeasible with value α—that is, if there is no degree d polynomial which gives an α-approximation to f—then the feasibility of the dual of this program will give us a "witness" to this fact. We refer to this witness as a dual polynomial for f. It is this witness that we will use to construct a tensor Q which witnesses that μ^α is large.

Lemma 19

$$\alpha_d(f) = \max_v \left\{ \frac{1 + \langle v, f \rangle}{1 - \langle v, f \rangle} : \|v\|_1 = 1, v^T W = 0 \right\}$$

Proof: Follows from duality theory of linear programming. □

Corollary 20 (Sherstov Corollary 3.3.1 [She08], Shi-Zhu Section 3.1 [SZ07]) *Let $f : \{-1,+1\}^n \to \mathbb{R}$ and let $d = \deg_\alpha(f)$. Then there exists a function $v : \{-1,+1\}^n \to \mathbb{R}$ such that*

1. *$\langle v, \chi_T \rangle = 0$ whenever $|T| \leq d$.*
2. *$\|v\|_1 = 1$.*
3. *$\langle v, f \rangle \geq \frac{\alpha-1}{\alpha+1}$.*

When $\alpha = \infty$, there is a function $v : \{-1,+1\}^n \to \mathbb{R}$ satisfying items (1), (2), and such that $v(x)f(x) \geq 0$ for all $x \in \{-1,+1\}^n$.

Špalek [Špa08] has given an explicit construction of an optimal dual polynomial for the OR function. For our analysis, however, we only make use of the properties guaranteed by Corollary 20.

5.2 Pattern Tensors

We define a natural generalization of the pattern matrices of Sherstov [She07] to the tensor case. We use a slightly different definition of pattern tensors than that of Chattopadhyay [Cha07] to allow the reduction to disjointness.

Let $\phi : \{-1,+1\}^m \to \mathbb{R}$ be a function and M a natural number. We define a k-dimensional pattern tensor $A_{k,M,\phi}$ as follows. Let $x \in \{-1,+1\}^{mM^{k-1}}$. We view $x = (x^1, \ldots, x^m)$ as consisting of m many blocks, where each $x_i \in \{-1,+1\}^{M^{k-1}}$ can be viewed as a $k-1$ dimensional tensor of side length M. We further let $y_i \in [M]^m$ for each $i = 1, \ldots, k-1$ and view each $y_i = (y_i[1], \ldots, y_i[m])$ as consisting of m-blocks where $y_i[j] \in [M]$ is an index into a side of x^i. Now define

$$A_{k,m,\phi}[x, y_1, \ldots, y_{k-1}] = \phi(x^1[y_1[1], \ldots, y_{k-1}[1]], \ldots, x^m[y_1[m], \ldots, y_{k-1}[m]]).$$

Note that $\mathrm{size}(A_{k,M,\phi}) = 2^{mM^{k-1}} M^{m(k-1)}$. We will often use the abbreviation $\bar{y} = (y_1, \ldots, y_{k-1})$. A nice property of pattern tensors is that every m-bit string z appears as input to ϕ an equal number of times, over all choices of $x, \bar{y}$.

The key lemma about pattern tensors is given next. Such a lemma was first shown by Chattopadhyay [Cha07]. Chattopadhyay and Ada [CA08] also show a statement similar to this one.

Lemma 21 *Let A be a $(k, M, c\cdot\phi)$ pattern tensor, where $c = 2^m \mathrm{size}(A)^{-1}$. Suppose that ϕ satisfies $\ell_1(\phi) = 1$ and $\hat{\phi}_T = 0$ for all sets $T \subseteq [m]$ with $|T| \leq d$. Then*

$$\mu^*(A) \leq 2^{-d}$$

provided that $M \geq 2e(k-1)2^{2^{k-1}}m/d$.

Proof: The idea of the proof will be to bound $\mathbb{E}[|A \bullet_1 A|]$ and apply Lemma 13 to obtain an upper bound on $\mu^*(A)$. For a string $\ell \in \{0,1\}^{k-1}$ we use the abbreviation $\bar{y}^\ell = (y_1^{\ell_1}, \ldots, y_{k-1}^{\ell_{k-1}})$. In particular, $\bar{y}^0 = (y_1^0, \ldots, y_{k-1}^0)$ and $\bar{y}^1 = (y_1^1, \ldots, y_{k-1}^1)$.

$$\mathbb{E}[|A \bullet_1 A|] = \left(\frac{2^m}{\text{size}(A)}\right)^{2^{k-1}} \mathbb{E}_{\bar{y}^0,\bar{y}^1} \left| \mathbb{E}_x \prod_{\ell=0}^{2^{k-1}-1} \sum_{T \subseteq [m]} \hat{\phi}(T) \prod_{i \in T} x^i[y_1^{\ell_1}[i], \ldots, y_{k-1}^{\ell_{k-1}}[i]] \right| \tag{4}$$

$$\leq \frac{1}{\text{size}(A)^{2^{k-1}}} \mathbb{E}_{\bar{y}^0,\bar{y}^1} \sum_{\substack{T_0,\ldots,T_{2^{k-1}-1} \\ |T_\ell| > d}} \prod_{i=1}^{m} \left| \mathbb{E}_{x^i} \prod_{\substack{\ell \in \{0,1\}^{k-1} \\ i \in T_\ell}} x^i[y_1^{\ell_1}[i], \ldots, y_{k-1}^{\ell_{k-1}}[i]] \right|. \tag{5}$$

Here we have used the fact that $\hat{\phi}(T) \leq 2^{-n} \ell_1(\phi) = 2^{-n}$.

We now develop a sufficient condition in terms of $\bar{y}^0, \bar{y}^1$ and $T_0, \ldots, T_{2^{k-1}-1}$, for the product of expectations over x^i to be zero. We say that $\bar{y}^0, \bar{y}^1$ select a nondegenerate cube in position i if $y_j^0[i] \neq y_j^1[i]$ for all $j = 1, \ldots, k-1$. The reason for this terminology is that in this case $(y_1^{\ell_1}[i], \ldots, y_{k-1}^{\ell_{k-1}}[i])$ define 2^{k-1} distinct points over $\ell \in \{0,1\}^{k-1}$. If this is not the case, we say that $\bar{y}^0, \bar{y}^1$ select a degenerate cube in position i.

Notice that if $\bar{y}^0, \bar{y}^1$ select a nondegenerate cube in position $i \in [m]$ and $i \in T_\ell$ for some $\ell \in \{0,1\}^{k-1}$ then

$$\mathbb{E}_{x^i} \prod_{\substack{\ell \in \{0,1\}^{k-1} \\ i \in T_\ell}} x^i[y_1^{\ell_1}[i], \ldots, y_{k-1}^{\ell_{k-1}}[i]] = 0.$$

We will now upper bound the probability over the choice of $\bar{y}^0, \bar{y}^1$ and $T_0, \ldots, T_{2^{k-1}-1}$ that this does not happen. Suppose that $\bar{y}^0, \bar{y}^1$ select g many degenerate cubes. By the above reasoning the number of sets $T_0, \ldots, T_{2^{k-1}-1}$ which lead to a nonzero expectation is at most

$$\left(\sum_{r=d+1}^{g} \binom{g}{r}\right)^{2^{k-1}} \leq 2^{g 2^{k-1}}.$$

Now we bound the probability that $\bar{y}^0, \bar{y}^1$ select g many degenerate cubes. The probability that $y_j^0[i] = y_j^1[i]$ is $1/M$. Thus by a union bound, the probability that a single cube is degenerate is at most $(k-1)/M$. Finally, as each index is chosen independently, the probability of g many degenerate cubes is at most

$$\binom{m}{g} \left(\frac{k-1}{M}\right)^g.$$

Putting everything together we have

$$
\begin{aligned}
\mathbb{E}[|A \bullet_1 A|] &\leq \frac{1}{\mathrm{size}(A)^{2^{k-1}}} \sum_{g=d+1}^{m} \binom{m}{g} \left(\frac{k-1}{M}\right)^g 2^{g 2^{k-1}} \\
&\leq \frac{1}{\mathrm{size}(A)^{2^{k-1}}} \sum_{g=d+1}^{m} \left(\frac{e(k-1)2^{2^{k-1}} m}{dM}\right)^g \\
&\leq \frac{2^{-d}}{\mathrm{size}(A)^{2^{k-1}}}
\end{aligned}
$$

provided that $M \geq 2e(k-1)2^{2^{k-1}} m/d$. □

Remark 22 *Our analysis cannot be improved by much without using more explicit information about the Fourier coefficients $\hat{q}(T)$ than given in Corollary 20. Apart from removing the Fourier coefficients, the only inequality we have used to arrive at Equation (5) is to turn an absolute value of a sum into a sum of absolute values. When $\bar{y}^0, \bar{y}^1$ select a degenerate cube, the most likely case is that it is what we call 1-degenerate—that is $y_i^0[t] = y_i^1[t]$ for* exactly one $1 \leq i \leq k-1$. *If the degenerate cubes selected by $\bar{y}^0, \bar{y}^1$ are all 1-degenerate, then one can see that the only sets $\{T_\ell\}$ which lead to a nonzero expectation are ones where the sets T_ℓ come in pairs. The number of such paired sets $\{T_\ell\}$ is not significantly smaller than the upper bound we give; furthermore, in this case all Fourier coefficients will be taken to an even power and so no cancellation occurs and the absolute value of the sum will be equal to the sum of absolute values.*

With this lemma in hand, we can now show our main result, proving a lower bound on $\mu^\alpha(A_{k,M,f})$ in terms of the approximate degree of f.

Theorem 23 *For a nonnegative integer m and a Boolean function f on m variables, and an integer $k \geq 2$*

$$\log \mu^\alpha(A_{k,M,f}) \geq \deg_{\alpha_0}(f)/2^{k-1} + \log \frac{\alpha_0 - \alpha}{\alpha_0 + 1},$$

for every $1 \leq \alpha < \alpha_0 < \infty$, provided $M \geq 2e(k-1)2^{2^{k-1}} m/\deg_{\alpha_0}(f)$.
Furthermore,

$$\log \mu^\infty(A_{k,M,f}) \geq \deg_\infty(f)/2^{k-1},$$

provided $M \geq 2e(k-1)2^{2^{k-1}} m/\deg_\infty(f)$

Proof: For simplicity we will drop the subscripts and just write A for $A_{k,M,f}$. Recall that

$$
\begin{aligned}
\mu^\alpha(A) &= \max_{Q:\|Q\|_1=1} \frac{(1+\alpha)\langle A, Q\rangle + (1-\alpha)}{2\mu^*(Q)} \\
\mu^\infty(A) &= \max_{Q:Q\circ A \geq 0} \frac{\langle A, Q\rangle}{\mu^*(Q)}.
\end{aligned}
$$

Let q be the vector from Corollary 20 which witnesses that the α_0-approximate degree of f is at least d. We let Q be the $(k, M, c \cdot q)$ pattern tensor where $c = 2^m/\text{size}(A)$. This choice of normalization implies that $\|Q\|_1 = 1$ as $\|q\|_1 = 1$.

First consider the case $1 \leq \alpha < \infty$. Then we have $\langle q, f\rangle \geq (\alpha_0 - 1)/(\alpha_0 + 1)$, and so $\langle A, Q\rangle \geq (\alpha_0 - 1)/(\alpha_0 + 1)$. This allows us to bound $(1/2)$ the term in the numerator of $\mu^\alpha(A)$ as follows:

$$\frac{(1+\alpha)\langle A, Q\rangle + (1-\alpha)}{2} \geq \frac{\alpha_0 - \alpha}{\alpha_0 + 1}.$$

In the case $\alpha = \infty$, observe that Q inherits the property $Q \circ A \geq 0$ as $q \circ f \geq 0$. The fact that $q \circ f \geq 0$ together with $\|q\|_1 = 1$ gives $\langle f, q\rangle = 1$, which in turn implies $\langle A, Q\rangle = 1$.

Let $d = \deg_{\alpha_0}(f)$ or $d = \deg_\infty(f)$, respectively. As q has no nonzero Fourier coefficients of degree less than d by Corollary 20, we can apply Lemma 21 to give

$$\mu^*(Q) \leq 2^{-d},$$

under the assumption that $M \geq 2e(k-1)2^{2^{k-1}}m/d$. The statement now follows from Lemma 13.
□

6 Applications

In this section, we apply Theorem 23 to prove lower bounds on the k-party number-on-the-forehead randomized communication complexity of disjointness. Then we formally state the implications this result has for proof systems via the results of Beame, Pitassi, and Segerlind [BPS06].

6.1 A lower bound for disjointness

Let $\text{OR}_n : \{-1,+1\}^n \to \{-1,+1\}$ be the OR function on n bits, and let $\text{DISJ}_{k,n} : (\{-1,+1\}^n)^k \to \{-1,+1\}$ be defined as $\text{DISJ}_{k,n}(x_1, \ldots, x_k) = -\text{OR}_n(x_1 \wedge x_2 \ldots \wedge x_k)$.

By embedding a pattern tensor into the tensor $\text{DISJ}_{k,n}$, we can get the following lower bound.

Corollary 24

$$R_{1/4}(\text{DISJ}_{k,n}) = \Omega\left(\frac{n^{1/(k+1)}}{2^{2^k}}\right)$$

Proof: The idea of the proof will be to embed an appropriate pattern tensor into $\text{DISJ}_{k,n}$ and apply Theorem 23. Let $c_k = 5e(k-1)2^{2^{k-1}}$. As Nisan and Szegedy have shown $\deg_3(\text{OR}_n) \geq \sqrt{n/6}$, we wish to define integers m, M such that $M \geq c_k\sqrt{m}$ and $mM^{k-1} \leq n$. To this end, let $m = \lfloor \frac{n}{(2c_k)^{k-1}} \rfloor$ and $M = c_k \lceil \sqrt{m} \rceil$. Let $n' = mM^{k-1}$. One can easily check that $n' \leq n$.

We will now see that the pattern tensor (k, M, OR_m) is a subtensor of $\text{OR}_{n'}(x_1 \wedge \ldots \wedge x_k)$. This will then give the result by the obvious reduction to $\text{DISJ}_{k,n}$.

Let A be the (k, M, OR_m) pattern tensor. Recall that

$$A[x, y_1, \ldots, y_{k-1}] = \mathrm{OR}_m(x^1(y_1[1], \ldots, y_{k-1}[1]), \ldots, x^m[y_1[m], \ldots, y_{k-1}[m]]),$$

where each $y_j[i] \in [M]$, and x^j is a $k-1$ dimensional tensor of side length M. To each $y_j[i]$ we associate a $k-1$ tensor z_j^i of side length M, where $z_j^i[t_1, \ldots, t_{k-1}] = 1$ if and only if $t_j = y_j[i]$. In this way, $x^1[y_1[1], \ldots, y_{k-1}[1]] = \mathrm{OR}_{M^{k-1}}(x^1 \wedge z_1^1 \wedge \ldots \wedge z_{k-1}^1)$. Letting $z_j = (z_j^1, \ldots, z_j^m)$ we have

$$\mathrm{OR}_{n'}(x_1 \wedge z_1 \ldots \wedge z_{k-1}) = \mathrm{OR}_m(\mathrm{OR}_{M^{k-1}}(x_1^1 \wedge z_1^1 \wedge \ldots \wedge z_{k-1}^1), \ldots, \mathrm{OR}_{M^{k-1}}(x_1^m \wedge z_1^m \wedge \ldots \wedge z_{k-1}^m)).$$

This shows that A is a subtensor of $-\mathrm{DISJ}_{k,n'}$. The result now follows from Theorem 23 and Theorem 7. □

Remark 25 *Note that a statement similar to that of Corollary 24 can be proved for any symmetric function, not just* OR*. But for some functions (e.g. threshold functions with threshold a constant fraction of* n*) much better bounds can be proved by reduction to inner product. For this reason, we do not include the general statement here.*

6.2 Proof systems

In this section we formally define the proof systems discussed in the introduction, and the lower bounds which follow from our results on disjointness.

A k-threshold formula is a formula of the form $\sum_j \gamma_j m_j \geq t$, where t, γ_j are integers, and each m_j is a monomial over variables $x_1, \ldots, x_n$. The size of a k-threshold formula is the sum of the sizes of γ_j and t, written in binary. For k-threshold formulas f_1, f_2, g, we say that g is *semantically entailed* by f_1 and f_2 if every 0/1 assignment to $x_1, \ldots, x_n$ that satisfies both f_1 and f_2 also satisfies g.

Let ϕ be an unsatisfiable CNF formula with variables $x_1, \ldots, x_n$. For each clause of ϕ we create a linear threshold formula which is satisfied if and only if the clause is. We refer to these clauses as *axioms*. We say that $\mathcal{P}$ is a $\mathrm{Th}(k)$ refutation of ϕ if

- $\mathcal{P}$ is a sequence $L_1, \ldots, L_t$ of k-threshold formulas.
- Each formula L_j is either an axiom or is semantically entailed by formulas $L_i, L_{i'}$ with $i, i' < j$.
- The final formula L_t is $0 \geq 1$.

The size of $\mathcal{P}$ is the sum of the sizes of $L_1, \ldots, L_t$. We say that $\mathcal{P}$ is *tree-like* if the underlying directed acyclic graph representing the implication structure of the proof is a tree.

We are now ready to state the connection of [BPS06] between the number-on-the-forehead complexity of disjointness and the size of $\mathrm{Th}(k)$ proofs.

Theorem 26 (Beame, Pitassi, and Segerlind [BPS06]) *Let $k \geq 2$ be a constant. For every n, there is a CNF formula ϕ on n variables such that the size of any $\mathrm{Th}(k-1)$ refutation of ϕ is at least*

$$\exp\left(\Omega\left(\frac{R^k_{1/4}(\mathrm{DISJ}_{k,m})}{\log n}\right)^{1/3}\right).$$

where $m = \frac{n^{2/3}}{2\log n}$.

Substituting the bounds from Corollary 24 we obtain the following.

Corollary 27 *Let $k \geq 2$ be a constant. For every n there is a CNF formula ϕ over n variables which requires $\mathrm{Th}(k-1)$ refutation proofs of size*

$$\exp\left(\Omega\left(\frac{n^{2/(9k+9)}}{(\log n)^{4/9}\, 2^{2^k/3}}\right)\right).$$

Acknowledgments

We greatly benefited during the course of this work from comments and conversations with Paul Beame, Harry Buhrman, Mike Saks, Gideon Schechtman, Nate Segerlind, Sasha Sherstov, Robert Špalek, Emanuele Viola, Fengming Wang, Avi Wigderson, and Ronald de Wolf. We would also like to thank the anonymous referees for their many suggestions.

References

[AA05] S. Aaronson and A. Ambainis. Quantum search of spatial regions. *Theory of Computing*, 1:47–79, 2005.

[Abe78] H. Abelson. Lower bounds on information transfer in distributed computations. In *Proceedings of the 19th IEEE Symposium on Foundations of Computer Science*, pages 151–158. IEEE, 1978.

[BHK01] L. Babai, T. Hayes, and P. Kimmel. The cost of the missing bit: communication complexity with help. *Combinatorica*, 21:455–488, 2001.

[BHN08] P. Beame and D. Huynh-Ngoc. Multiparty communication complexity of AC^0. Technical Report TR-08-082, ECCC, 2008.

[BNS89] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols and Logspace-hard pseudorandom sequences. In *Proceedings of the 21st ACM Symposium on the Theory of Computing*, pages 1–11. ACM, 1989.

[BPS06] P. Beame, T. Pitassi, and N. Segerlind. Lower bounds for Lovász-Schrijver systems and beyond follow from multiparty communication complexity. *SIAM Journal on Computing*, 37(3):845–869, 2006.

[BPSW06] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A strong direct product lemma for corruption and the NOF complexity of disjointness. *Computational Complexity*, 15(4):391–432, 2006.

[BT94] R. Beigel and J. Tarui. On ACC. *Computational Complexity*, 4:350–366, 1994.

[CA08] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. Technical Report TR-08-002, ECCC, 2008.

[CFL83] A. Chandra, M. Furst, and R. Lipton. Multi-party protocols. In *Proceedings of the 15th ACM Symposium on the Theory of Computing*, pages 94–99. ACM, 1983.

[Cha07] A. Chattopadhyay. Discrepancy and the power of bottom fan-in depth-three circuits. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 449–458. IEEE, 2007.

[Chu90] F. Chung. Quasi-random classes of hypergraphs. *Random Structures and Algorithms*, 1:363–382, 1990.

[DPV08] M. David, T. Pitassi, and E. Viola. Improved separations between nondeterministic and randomized multiparty communication. In *APPROX-RANDOM*, volume 5171 of *Lecture Notes in Computer Science*, pages 371–384. Springer, 2008.

[FG05] J. Ford and A. Gál. Hadamard tensors and lower bounds on multiparty communication complexity. In *Proceedings of the 32th International Colloquium On Automata, Languages and Programming*, pages 1163–1175, 2005.

[Gro94] V. Grolmusz. The BNS lower bound for multi-party protocols is nearly optimal. *Information and computation*, 112(1):51–54, 1994.

[GW95] M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995.

[HG91] J. Håstad and M. Goldmann. On the power of small-depth threshold circuits. *Computational Complexity*, 1:113–129, 1991.

[IK06] D. Itsykson and A. Kojevnikov. Lower bounds of static Lovász-Schrijver calculus proofs for Tseitin tautologies. *Zapiski Nauchnyh Seminarov POMI*, 340:10–32, 2006.

[Jam87] G. J. O. Jameson. *Summing and nuclear norms in banach space theory*. Cambridge University Press, 1987.

[KKN95] M. Karchmer, E. Kushilevitz, and N. Nisan. Fractional covers and communication complexity. *SIAM Journal on Discrete Mathematics*, 8(1):76–92, 1995.

[Kla01] H. Klauck. Lower bounds for quantum communication complexity. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*. IEEE, 2001.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[KS87] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. In *Proceedings of the 2nd Annual Conference on Structure in Complexity Theory*, pages 41–49, 1987.

[LMSS07] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007.

[LS91] L. Lovász and A. Schrijver. Cones of matrices and set-functions, and 0-1 optimization. *SIAM Journal of Optimization*, 1:1–17, 1991.

[LS07] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. In *Proceedings of the 39th ACM Symposium on the Theory of Computing*, pages 699–708. ACM, 2007.

[Raz00] R. Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

[Raz03] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003.

[She07] A. Sherstov. Separating AC^0 from depth-2 majority circuits. In *Proceedings of the 39th ACM Symposium on the Theory of Computing*, pages 294–301. ACM, 2007.

[She08] A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Proceedings of the 40th ACM Symposium on the Theory of Computing*, pages 85–94. ACM, 2008.

[Špa08] R. Špalek. A dual polynomial for OR. Technical Report arXiv:0803.4516 [cs.CC], arXiv, 2008.

[SZ07] Y. Shi and Y. Zhu. Quantum communication complexity of block-composed functions. Technical Report arXiv:0710.0095 [quant-ph], arXiv, 2007.

[Tes02] P. Tesson. *Communication complexity questions related to finite monoids and semigroups*. PhD thesis, McGill University, 2002.

[VW07] E. Viola and A. Wigderson. One-way multi-party communication lower bound for pointer jumping with applications. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*. IEEE, 2007.

[Yao79] A. Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th ACM Symposium on the Theory of Computing*, pages 209–213. ACM, 1979.

[Yao83] A. Yao. Lower bounds by probabilistic arguments. In *Proceedings of the 24th IEEE Symposium on Foundations of Computer Science*, pages 420–428, 1983.